

Implementation of Data Mining to Predict the Feasibility of Blood Donors Using C4.5 Algorithm

¹Anita Febriani, ²Tiara Trimadya Rahmawati, ³Eka Sabna

^{1,2,3}Program Studi Teknik Informatika, STMIK Hang Tuah Pekanbaru

Email: ¹nitasuheri@gmail.com, ²tiaratrimadyawan@gmail.com, ³es3jelita@yahoo.com

Article Info

Article history:

Received Feb3th, 2018

Revised Feb21th, 2018

Accepted Mar10th, 2018

Keywords:

Blood donor

C4.5

Decision tree

Feasibility

Prediction

ABSTRACT

Blood Transfusion Unit PMI Pekanbaru City is part of a company or agency that serves blood donation, every blood bag obtained from the community voluntarily come to PMI to donate blood with the goal of humanity. In Blood Transfusion Unit PMI Pekanbaru City, has a provision to be blood donors that must be met in order to donate blood in UTD PMI Pekanbaru City. Data Mining is a combination of a number of computer science disciplines that are defined as the process of discovering new patterns from massive data sets. By using RapidMiner software and using the method of Decision Tree Algorithm C4.5 to determine the eligibility of blood donors based on Age, Weight, Hemoglobin, and Blood Pressure. In the study of hemoglobin is the most decisive variable in blood donors. And the result accuracy is 94,02% which means the accuracy of this model is very good.

Copyright © 2018 Puzzle Research of Data Technology

Corresponding Author:

Anita Febriani

Program Studi Teknik Informatika

STMIK Hang Tuah Pekanbaru

Jl. Mustafa Sari No . 05 Tangkerang Selatan, Pekanbaru, Riau

Email: nitasuheri@gmail.com

1. PENDAHULUAN

Unit Transfusi Darah (UTD) Palang Merah Indonesia (PMI) Kota Pekanbaru merupakan bagian perusahaan atau instansi yang melayani pendonoran darah, setiap kantong darah yang diperoleh dari masyarakat dengan sukarela datang ke PMI untuk mendonorkan darah dengan tujuan kemanusiaan. Unit ini memiliki ketentuan-ketentuan untuk menjadi pendonor darah yang harus dipenuhi agar bisa mendonorkan darahnya di UTD PMI Kota Pekanbaru. Seringnya pihak PMI mengadakan donor darah di berbagai tempat, semakin banyak lembaran formulir dari calon pendonor dan juga menumpuk. Lembaran formulir tersebut kemudian diarsipkan setiap bulannya, tanpa ada analisa lebih lanjut. Agar data yang telah diarsipkan tersebut dapat memberikan informasi yang bermanfaat maka perlu dilakukan analisa. Berdasarkan uraian tersebut, penelitian ini menggunakan data mining untuk memprediksi kelayakan calon pendonor darah berdasarkan usia, berat badan, kadar hemoglobin, tekanan darah (sistolik dan diastolik) dengan memanfaatkan tumpukan data yang telah dimiliki UTD PMI Kota Pekanbaru. Melalui data pendonor yang ada pada PMI dapat dilakukan proses mining untuk menentukan pola prediksi pendonor yang berpotensi menjadi pendonor tetap. Sehingga dapat difokuskan dimana penyebaran informasi harus dilakukan agar lebih efisien dalam menambah jumlah pendonor darah.

Penelitian ini menggunakan metode Algoritma C4.5 dalam memprediksi seseorang apakah bisa mendonorkan darahnya atau tidak berdasarkan: usia, berat badan, kadar hemoglobin (HB), tekanan darah (sistolik dan diastolik) yang dimilikinya sebagai variabel pendukung. Data usia, berat badan, kadar Hemoglobin, dan tekanan darah, hasilnya diambil dari beberapa literatur tentang usia, berat badan, kadar Hemoglobin, dan tekanan darah, baik menurut World Health Organization (WHO), Joint National Committee 7 (JNC 7), dan Departemen Kesehatan Republik Indonesia [1]. Menurut Han dan Kamber [2] klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Algoritma C4.5 dapat menangani data numerik dan diskret. Algoritma C4.5

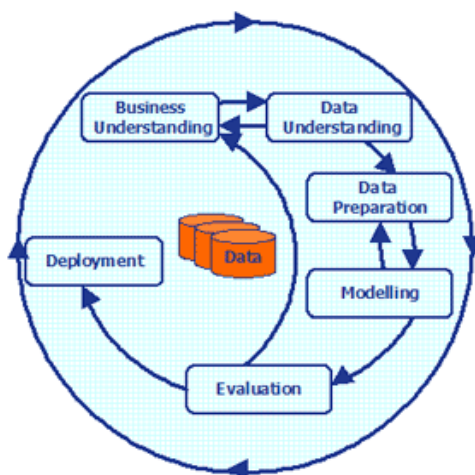
menggunakan rasio perolehan (gain ratio). Sebelum menghitung rasio perolehan, perlu dilakukan perhitungan nilai informasi dalam satuan bits dari suatu kumpulan objek, yaitu dengan menggunakan konsep entropy.

Algoritma C4.5 akan menghasilkan pohon keputusan. Pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Dengan pohon keputusan, manusia dapat dengan mudah mengidentifikasi dan melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah sehingga dengan memperhitungkan faktor-faktor tersebut dapat dihasilkan penyelesaian terbaik untuk masalah tersebut. Pohon keputusan ini juga dapat menganalisa nilai resiko dan nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah.

Penelitian oleh Atmaja [2] dari hasil analisis yang dilakukan, dapat diketahui bahwa data yang digunakan adalah data random sebanyak 600 data, dimana data training berjumlah 500 data, dan data set berjumlah 100 data. Metode yang dipakai dalam melakukan data mining adalah metode klasifikasi menggunakan algoritma C4.5. Dari pohon decision tree yang didapatkan dari proses data mining yang dilakukan disimpulkan bahwa pegawai swasta dengan umur diatas 26 tahun paling banyak menjadi pendonor. Penelitian Kusuma [5] dari hasil penelitiannya dapat meminimalisasi dan mengurangi tingkat kerumitan perhitungan dari proses penginputan data penentuan calon pendonor darah.

2. METODE PENELITIAN

Metode yang diterapkan dalam penelitian adalah metode Cross Industry Standard Proses for Data Mining (CRISP-DM) merupakan suatu metodologi data mining yang disusun oleh konsorsium perusahaan yang didirikan oleh Komisi Eropa pada tahun 1996 [4] dan telah ditetapkan sebagai proses standar dalam data mining. Menurut Larose, data mining memiliki enam fase CRISP-DM, seperti yang tertera pada Gambar 1 [3].



Gambar 1. Tahapan Metode CRISP-DM [3]

2.1. Fase Pemahaman Bisnis (*Business Understanding Phase*)

Pada fase ini peneliti memahami tujuan proyek (penelitian) dan kebutuhan tujuan bisnis (penelitian). Tujuan proyek (penelitian) untuk mengetahui kelayakan pada calon pendonor darah. Kebutuhan tujuan bisnis (penelitian) ini untuk mengetahui berdasarkan variabel usia, berat badan, kadar Hemoglobin, dan tekanan darah, mana yang lebih menentukan untuk calon pendonor dapat mendonorkan darahnya. Dan setelah mengetahui tujuan proyek ini, selanjutnya melakukan proses menghilangkan noise dan inkonsisten pada data.

2.2. Fase Pemahaman Data (*Data Understanding Phase*)

Pada fase ini peneliti membuat target data dan fokus pada variabel atau sampel data yang akan diambil. Setelah itu sampel data di cleaning target data dengan tujuan untuk mendapatkan data yang konsisten. Data yang ada di dalam database tidak semuanya digunakan, oleh karena itu hanya data yang sesuai untuk di analisis yang akan diambil.

2.3. Fase Pengolahan Data (*Data Preparation Phase*)

Setelah melakukan fase pemahaman data, maka pada fase pengolahan data ini membuat dataset final yang akan diterapkan ke dalam alat pemodelan. Dari data mentah awal dan selanjutnya akan melakukan proses data mining. Data diubah atau digabung ke dalam format yang sesuai untuk diproses di dalam data mining. Data tersebut adalah dengan menggunakan nama pendonor di UTD PMI KOTA PEKANBARU dan

karena aplikasi yang digunakan adalah *RapidMiner*, jadi nama pendonor tidak perlu diubah ke dalam bentuk nomor, karena *RapidMiner* sebelum menampilkan hasil dapat memilih tabel yang akan di keluarkan untuk menjadi akar (*tree*).

2.4. Fase Pemodelan (*Modeling Phase*)

Pada fase pemodelan ini peneliti melakukan pemilihan teknik klasifikasi dengan menggunakan algoritma C4.5. Fase ini merupakan proses utama karena saat metode diterapkan maka didapat pengetahuan yang tersembunyi dari data.

2.5. Fase Evaluasi (*Evaluation Phase*)

Fase Evaluasi ini akan mengavaluasi dan meneliti untuk menyakinkan kalau tahap pemodelan yang digunakan memenuhi tujuan dari penelitian. Yakni fase untuk menarik pola-pola ke dalam pengetahuan yang ditemukan. Fase ini merupakan evaluasi untuk menilai apakah model yang telah dibuat sesuai dengan tujuan penelitian. Fase ini untuk mengetahui keakurasian dari model menggunakan *confision matrix*.

2.6. Fase Penyebaran (*Deployment Phase*)

Deployment merupakan tahapan akhir dalam pembuatan laporan hasil kegiatan data mining. Laporan akhir yang berisi mengenai pengetahuan yang diperoleh atau pengenalan pola pada data dalam proses data mining.

2.7. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (Decision Tree). Pohon keputusan merupakan metode klasifikasi dan prediksi yang terkenal [5]. Pohon keputusan berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target [6]. Pemilihan atribut pada algoritma induksi decision tree menggunakan ukuran berdasarkan entropy yang dikenal dengan information gain sebagai sebuah heuristic untuk memilih atribut yang merupakan bagian terbaik dari contoh ke dalam kelas. Semua atribut adalah bersifat kategori yang bernilai diskrit. Atribut dengan nilai continuous harus didiskritkan [7]. Perhitungan nilai entropy dan nilai gain dapat ditunjukkan pada persamaan 1 dan 2 berikut [8]:

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan:

- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi S
- pi : proporsi dari S_i terhadap S

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

- S : Himpunan kasus
- A : Fitur
- n : Jumlah partisi A
- $|S_i|$: Jumlah kasus pada partisi ke-i
- $|S|$: Jumlah kasus dalam S

2.8. Donor Darah

Penyumbang darah atau Donor darah adalah proses pengambilan darah dari seseorang secara sukarela atau pengganti untuk disimpan di bank darah sebagai stok darah untuk kemudian digunakan untuk transfusi darah [9][10]. Beberapa keuntungan dari donor darah adalah: (1) Menurunkan risiko terkena penyakit jantung dan pembuluh darah, (2) Menurunkan risiko kanker, (3) Membantu menurunkan berat badan, (4) Mendeteksi penyakit serius, dan (5) Membuat lebih sehat secara psikologis dan memperpanjang usia [11][12].

3. HASIL DAN PEMBAHASAN

3.1. Pengolahan Data

Dalam penelitian ini data yang digunakan adalah data calon pendonor darah dari UTD PMI Kota Pekanbaru sebanyak 1839 data. Data sampel pendonor darah dapat dilihat pada Tabel 1. Tabel 2 adalah pengelompokan dari pembagian dari masing-masing variabel.

Tabel 1. Data sampel dari data calon pendonor darah

No	Usia	Berat Badan	Hemoglobin	Tekanan Darah
1	Dewasa I	Gemuk I	Tinggi	Rendah
2	Lansia	Gemuk II	Tinggi	Tinggi
3	Lansia	Gemuk II	Normal	Tinggi
4	Lansia	Kurus	Tinggi	Normal
5	Dewasa II	Gemuk I	Tinggi	Rendah
6	Lansia	Gemuk II	Tinggi	Normal
7	Dewasa II	Sedang	Tinggi	Rendah
8	Dewasa I	Gemuk I	Tinggi	Tinggi
9	Lansia	Gemuk II	Rendah	Normal
10	Dewasa I	Kurus	Normal	Rendah
....				
1839	Remaja	Kurus	Normal	Rendah

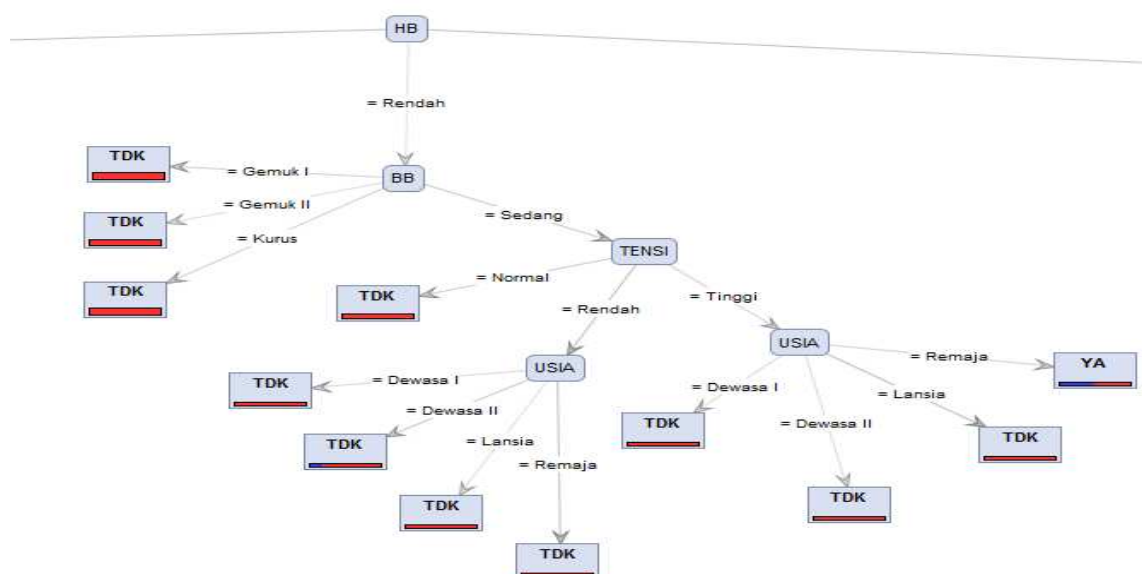
Tabel 2. Pembagian variabel dan kelas data

Nama Field	Pembagian Variabel dan Kelas Data
Usia	Remaja, Dewasa I, Dewasa II, Lansia
Berat Badan	Kurus, Sedang, Gemuk I, Gemuk II
Kadar <i>Hemoglobin</i>	Rendah, Normal, Tinggi
Tekanan Darah	Rendah, Normal, Tinggi

Atribut usia, berat badan, kadar *hemoglobin*, dan tekanan darah berdasarkan referensi dari WHO, JNC 7, dan DEPKES RI. Kemudian dirangkum dan disesuaikan dengan data yang dimiliki.

3.2. Pemodelan

Pemodelan dilakukan menggunakan *software* RapidMiner. Hasil dari *software RapidMiner* dalam bentuk pohon keputusan terlihat variabel *hemoglobin* yang menjadi node akar. Dan *hemoglobin* memiliki 3 nilai, yaitu 'rendah', 'normal', dan 'tinggi'. Gambar 2 menampilkan *hemoglobin* dengan dengan nilai 'rendah'.



Gambar 2. Pohon keputusan bagian nilai *hemoglobin* = 'rendah'

3.3. Evaluasi

Berdasarkan *confision matrix* pada Tabel 3, didapat akurasi dari pohon keputusan yang dihasilkan dari C4.5 untuk penentuan kelayakan calon pendonor darah adalah 94,02%. Hal ini berarti akurasi dari model ini dinyatakan sangat baik.

Tabel 3. Hasil evaluasi

		Observasi	
		Ya	Tidak
Prediksi	Ya	495	32
	Tidak	1	24

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah:

- (1) Atribut yang bisa digunakan untuk prediksi calon pendonor darah menggunakan Algoritma C4.5 adalah usia, berat badan, hemoglobin dan tekanan darah.
- (2) *Hemoglobin* merupakan atribut yang paling berpengaruh dalam penentuan kelayakan calon pendonor darah.
- (3) Akurasi pohon keputusan Algoritma C4.5 dalam penentuan kelayakan calon pendonor darah adalah 94,02%.

REFERENSI

- [1] PERMENKES NO 91 TAHUN 2015. 2016. "Standar Pelayanan Transfusi Darah". <http://peraturan.go.id/permen/kemenkes-nomor-91-tahun-2015-tahun-2016.html> diakses tanggal [Senin, 09-10-2017. Pukul 20.18].
- [2] Han, J dan Kamber, M. "Data Mining Concept and Techniques Second Edition". *Morgan Kauffman*. San Francisco. 2006.
- [3] Larose, Daniel T. "Discovering Knowledge in Data: An Introduction to Data Mining". John Willey & Sons, Inc. 2005.
- [4] Wirth R dan Hipp J, "CRISP-DM: Towards a Standard Process Model for Data Mining". *International Journal of Computing*. 2010.
- [5] Buulolo E, Silalahi N, Fadlina dan Rahim R, "C4.5 Algorithm to Predic the Impact of The Earthquake". *International Journal of Engineering Research and Technology*, 2017, Vol. 6, No. 2, pp: 10-15.
- [6] Ruggieri S, "Efficient C4.5 (classification algorithm)", *IEEE Transactions on Knowledge and Data Engineering*, 2012, Vol. 14, No. 2, pp: 438 – 444.
- [7] Hssina B, Merbouha A, Ezzikouri H dan Erritali M, "A Comparative study of decision tree ID3 and C4.5", *International Journal of Advanced Computer Science and Applications*, 2011, pp. 13-19.
- [8] Purnamasari M and Sulistiyono, "Decision Support System for Classification of Child Intelligence Using C4.5 Algorithm," *International Journal of Advanced Research in Computer Science*, 2014, vol. 5, pp. 16-20.
- [9] Palang Merah Indonesia, Donor Darah. diakses tanggal 20 September 2015
- [10] Kementerian Kesehatan RI, Pemanfaatan Kesehatan dalam Rangka Menumbuh Kembangkan Hidup Sehat Melalui Donor Darah, 2010.
- [11] Dewi RRR, Kartasurya MI dan Mawarni A, "Analisis Kebijakan Donor Darah Dan Implementasi Program Rekrutmen Donor Di Unit Donor Darah (UDD PMI) Kota Pontianak", *Jurnal Manajemen Kesehatan Indonesia*, 2016, Vol. 4, No. 2, pp:109-117.
- [12] Haspari NYD dan Herdiana I, "Hubungan antara Self-Esteem dengan Intensi Perilaku Prosocial Donor Darah pada Donor di Unit Donor Darah PMI Surabaya", *Jurnal Psikologi Kepribadian dan Sosial*, 2012, Vol. 1, No.03, pp: 173-182.

BIBLIOGRAFI PENULIS



Anita Febriani, saat ini penulis sebagai tenaga pengajar di Program Studi Teknik Informatika STMIK Hang Tuah Pekanbaru. Lulus S1 pada tahun 2011 di Jurusan Teknik Informatika UIN Sultan Syarif Kasim Riau. Lulus S2 pada tahun 2016 di Jurusan Magister Teknik Informatika Universitas Bina Nusantara. Bidang keahlian yang ditekuni yaitu Kecerdasan Buatan



Tiara Trimadya Rahmawati, yang biasa dipanggil Tiara. Lulus D3 Manajemen Informatika di Universitas Riau. Lulus S1 pada program studi teknik informatika STMIK Hang Tuah Pekanbaru pada tahun 2018.



Eka Sabna, Saat ini, penulis merupakan tenaga pengajar di Program Studi Teknik Informatika STMIK Hang Tuah Pekanbaru. Lulus S1 pada tahun 1994 di Universitas Putra Indonesia Yptk Padang. Lulus S2 Ilmu Pendidikan pada tahun 2004 di Universitas Negeri Padang. Lulus S2 Ilmu Komputer pada tahun 2006 di Universitas Putra Indonesia Yptk Padang. Bidang keahlian yang ditekuni yaitu Data Mining, Sistem Pakar, Logika Fuzzy